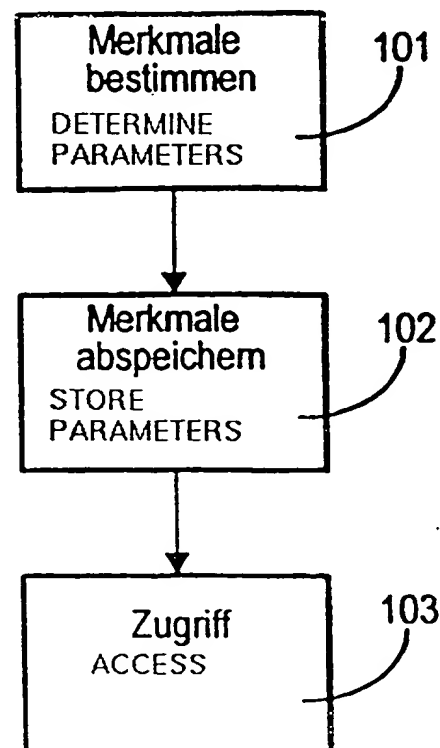


09/803.424

**PCT**WELTORGANISATION FÜR GEISTIGES EIGENTUM  
Internationales BüroINTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE  
INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

<b>(51) Internationale Patentklassifikation <sup>6</sup> :</b> <b>G11B 27/28, 27/10, G06F 17/30, G10L 5/06</b>	<b>A1</b>	<b>(11) Internationale Veröffentlichungsnummer:</b> <b>WO 99/05681</b>  <b>(43) Internationales Veröffentlichungsdatum:</b> 4. Februar 1999 (04.02.99)
<b>(21) Internationales Aktenzeichen:</b> PCT/DE98/01985 <b>(22) Internationales Anmeldedatum:</b> 15. Juli 1998 (15.07.98)  <b>(30) Prioritätsdaten:</b> 197 31 718.9      23. Juli 1997 (23.07.97)      DE  <b>(71) Anmelder (für alle Bestimmungsstaaten ausser US):</b> SIEMENS AKTIENGESELLSCHAFT [DE/DE]; Wittelsbacherplatz 2, D-80333 München (DE).  <b>(72) Erfinder; und</b> <b>(75) Erfinder/Anmelder (nur für US):</b> KAUP, Andre [DE/DE]; Brotmannstrasse 25, D-85635 Höhenkirchen (DE).		<b>(81) Bestimmungsstaaten:</b> CN, JP, KR, US, europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Veröffentlicht</b> <i>Mit internationalem Recherchenbericht.</i> <i>Vor Ablauf der für Änderungen der Ansprüche zugelassenen</i> <i>Frist; Veröffentlichung wird wiederholt falls Änderungen</i> <i>eintreffen.</i>
<b>(54) Title:</b> PROCESS FOR STORING SEARCH PARAMETERS OF AN IMAGE SEQUENCE AND ACCESS TO AN IMAGE STREAM IN SAID IMAGE SEQUENCE  <b>(54) Bezeichnung:</b> VERFAHREN ZUM ABSPEICHERN VON SUCHMERKMALEN EINER BILDSEQUENZ UND ZUGRIFF AUF EINE BILDFOLGE IN DER BILDSEQUENZ  <b>(57) Abstract</b>  The invention relates to a method for storing search parameters of an image sequence and accessing an image that is a true subset of the image sequence. Said image sequence contains audio information. In order to be able to selectively search within an image sequence containing audio information, search characteristics are determined from the audio information by means of voice recognition. A concept which is to be searched can be inputted by means of spoken language. Search characteristics are either single words of a predetermined language or phonemes/phoneme combinations. The inventive method thus enables image sequences to be automatically indexed with their audio information.  <b>(57) Zusammenfassung</b>  Verfahren zum Abspeichern von Suchmerkmalen einer Bildsequenz, die Toninformation umfaßt, und Zugriff auf eine Bildfolge, die eine echte Teilmenge der Bildsequenz ist. Um in einer Bildsequenz, die Toninformation umfaßt, gezielt suchen zu können, werden Suchmerkmale aus der Toninformation mittels eines Spracherkenners ermittelt. Auch die Eingabe eines zu suchenden Begriffs kann über gesprochene Sprache erfolgen. Suchmerkmale sind alternativ einzelne Wörter einer vorgegeben Sprache oder Phoneme/Phonemkombinationen. Mit diesem Verfahren können Bildsequenzen anhand ihrer Toninformation automatisch indiziert werden.		



# LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidshjan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	ML	Mali	TR	Türkei
BG	Bulgarien	HU	Ungarn	MN	Mongolei	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MR	Mauretanien	UA	Ukraine
BR	Brasilien	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Island	MX	Mexiko	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	NE	Niger	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NL	Niederlande	VN	Vietnam
CG	Kongo	KE	Kenia	NO	Norwegen	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NZ	Neuseeland	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	PL	Polen		
CM	Kamerun	KR	Republik Korea	PT	Portugal		
CN	China	KZ	Kasachstan	RO	Rumänien		
CU	Kuba	LC	St. Lucia	RU	Russische Föderation		
CZ	Tschechische Republik	LI	Liechtenstein	SD	Sudan		
DE	Deutschland	LK	Sri Lanka	SE	Schweden		
DK	Dänemark	LR	Liberia	SG	Singapur		
EE	Estland						

Beschreibung**Verfahren zum Abspeichern von Suchmerkmalen einer Bildsequenz und Zugriff auf eine Bildfolge in der Bildsequenz**

5

Die Erfindung betrifft ein Verfahren zum Abspeichern von Suchmerkmalen einer Bildsequenz, die Toninformation umfaßt, und einen Zugriff auf eine Bildfolge innerhalb der Bildsequenz anhand der Suchmerkmale.

10

Eine Bildsequenz ist eine Reihe zusammenhängender Bilder mit entsprechender Toninformation. Eine Bildfolge bezeichnet eine echte Teilmenge der Bildsequenz. Vorgebbare Information ist eine in den Suchmerkmalen zu suchende Information.

15

Vielfältige audiovisuelle Information (Bildsequenzen) wird heute in digitalisierter Form an verschiedenen Orten gespeichert. Um diese audiovisuelle Information zu nutzen, ist es nötig, die interessierenden Daten zunächst einmal aufzufinden.

20

Ein Spracherkennungssystem ist aus [1] bekannt.

25

Die MPEG-Standards zur Bildkompression sind dem Fachmann hinlänglich bekannt.

30

Die Aufgabe der Erfindung besteht darin, aus einer Bildsequenz Suchmerkmale zu bestimmen und abzuspeichern und somit einen Zugriff auf interessierende Daten zu gewährleisten.

Diese Aufgabe wird gemäß den Merkmalen der Patentansprüche 1 und 9 gelöst.

35

Es wird ein Verfahren zum Abspeichern von Suchmerkmalen einer Bildsequenz, die Toninformation umfaßt, dargestellt, bei dem die Suchmerkmale aus der Toninformation der Bildsequenz

ermittelt und abgespeichert werden. Weiterhin gibt eine vorgegebene Information an, welches Suchmerkmal in der Bildsequenz gefunden werden soll. Wird eine Übereinstimmung zwischen der vorgegebenen Information und den Suchmerkmalen der Bildsequenz ermittelt, so wird zu der Bildfolge, die mit dem gefundenen Suchmerkmal verknüpft ist, gesprungen.

Ansonsten, wird also keine Übereinstimmung zwischen der vorgegebenen Information und den Suchmerkmalen ermittelt, wird dem Benutzer eine entsprechende Rückmeldung angezeigt.

Befinden sich innerhalb der Bildsequenz mehrere Suchmerkmale, die auf die vorgegebene Information passen, so können die jeweilig verknüpften Bildfolgen einzeln der Reihe nach angesprungen werden oder eine Auswahl mit zusätzlichen zu den jeweiligen Bildfolgen gehörenden Suchmerkmalen in einer Übersicht ausgegeben werden. Hierzu sind verschiedene Möglichkeiten denkbar, die allgemein aus auf Datenbanken anwendbaren Suchtechniken bekannt sind.

Daraus ergibt sich der Vorteil einer automatisch für die jeweilige Bildsequenz resultierenden Menge von Suchmerkmalen, die einfach zu ermitteln sind und mit der Bildsequenz oder getrennt von der Bildsequenz in einer Datenbank abgespeichert werden können.

Eine Weiterbildung der Erfindung besteht darin, die Toninformation mittels eines Spracherkennungssystems auszuwerten im Hinblick auf vorgegebene Wortarten oder vorgegebene Worte. So kann eine vorgegebene Wortart alle Substantive umfassen, die mit der Bildsequenz oder getrennt von der Bildsequenz als Suchmerkmale abgespeichert werden.

Eine andere Weiterbildung besteht darin, die vorgegebene Information mittels gesprochener Sprache einzugeben. Hierzu kann vorteilhaft der Spracherkenner, der für die Erkennung der Suchmerkmale eingesetzt worden ist, verwendet werden. Die Eingabe von natürlich gesprochener Sprache hat u.a. den

Vorteil, daß ohne eine zur Verfügung stehende Tastatur oder andere Instrumentierung völlig ohne Benutzung z.B. der Hände die vorgegebene Information eingegeben werden kann.

Beispielsweise beim Führen eines Kraftfahrzeugs ist eine

- 5 Eingabe, bei der der Blick nicht vom Verkehrsgeschehen abgewandt werden muß, von Vorteil.

Auch ist es eine mögliche Weiterbildung, aus den mittels des Spracherkenners erkannten Suchmerkmalen eine Liste

- 10 anzubieten, die der Benutzer sowohl zur einfachen Auswahl eines Suchmerkmals, das er nicht aktiv kennen muß, benutzen kann. Ferner ist es möglich, eine Liste von beispielsweise Substantiven sortiert nach der Häufigkeit anzubieten, so daß der Benutzer z.B. die aktuellsten Bildfolgen über die
- 15 Suchmerkmale adressieren kann, wenn die Bildsequenz z.B. Nachrichtensendungen umfaßt.

Eine zusätzliche Weiterbildung besteht darin, als

- Suchmerkmale Phoneme und/oder Phonemkombination abzuspeichern
- 20 und somit den letzten Schritt der Spracherkennung, das Zuordnen zu real existierenden Worten, einzusparen. Dadurch ergibt sich ein flexibler Einsatz, da nicht für jede Sprache ein eigenes Lexikon mit den jeweiligen Umsetzungen in Phoneme und/oder Phonemkombinationen bereitstehen muß.

25

Ein anderes Verfahren ermöglicht direkt die Suche nach einer vorgegebenen Information innerhalb einer Toninformation umfassenden Bildsequenz, wobei die vorgegebene Information in der Bildsequenz gesucht wird und eine Bildfolge ab dem Bild

30 angezeigt wird, zu dessen Toninformation die vorgegebene Information zugeordnet werden kann. Dadurch können vorteilhaft, ohne Aufwand für Abspeichern und Verwalten von Suchmerkmalen, einzelne, vorzugsweise kurze, Bildsequenzen nach vorgegebener Information abgesucht werden.

35

Eine Weiterbildung besteht darin, mittels eines Spracherkennungssystems vorgebbare Worte aus der Toninformation zu erkennen.

- 5 Die Eingabe der vorgegebenen Information kann durch natürlich gesprochene Sprache erfolgen, die von dem Spracherkennungssystem erkannt wird.

Eine andere Weiterbildung ist die Suche nach einer  
10 Übereinstimmung zwischen der vorgegebenen Information und der Toninformation auf Basis von Phonemen und/oder Phonemkombinationen. Dabei ist es ein Vorteil, daß der letzte Schritt der Spracherkennung, die Zuordnung einer  
15 Phonemkombination zu einem Wort, eingespart wird, wobei damit keine sprachspezifischen Wörterbücher für diese Zuordnung im Spracherkenner vorhanden sein müssen. Es können auf diese Weise Laute und Lautkombinationen innerhalb der zu der Bildsequenz assoziierten Toninformation gefunden werden.

- 20 Weiterbildungen der Erfindung ergeben sich aus den abhängigen Ansprüchen.

Anhand der folgenden Figuren werden Ausführungsbeispiele der Erfindung näher dargestellt.

25

Es zeigen

- Fig.1 ein Blockdiagramm, das Schritte zur Durchführung  
eines Verfahrens zum Abspeichern einer Bildsequenz  
und Zugriff auf diese Bildsequenz enthält,  
30 Fig.2 eine Skizze, die eine Bildsequenz zeigt, die  
Bilddaten und Toninformation umfaßt,  
Fig.3 eine Skizze, die eine Möglichkeit zum Abspeichern von  
Suchmerkmalen zeigt,  
Fig.4 eine Skizze, die eine Aufteilung einer Liste aus  
35 Suchmerkmalen darstellt,  
Fig.5 eine Skizze, die das Zusammenspiel zwischen  
Toninformation, Spracherkenner, einem Mikrofon für

zusätzliche Spracheingabe und den Suchmerkmalen symbolisch veranschaulicht.

In Fig.1 ist ein Blockdiagramm dargestellt, das Schritte zur  
5 Durchführung eines Verfahrens zum Abspeichern einer  
Bildsequenz, die Toninformation umfaßt, und einen Zugriff auf  
diese Bildsequenz enthält.

Dazu werden in einem Schritt 101 Suchmerkmale aus der  
10 Toninformation ermittelt. Dies geschieht mittels eines  
Spracherkenners, der Worte oder Laute aus der Toninformation  
erkennt und diese in ihrer Gesamtheit oder nach vorgebbaren  
Merkmalen gefiltert (siehe unten) als Suchmerkmale  
abspeichert.

15 Die Suchmerkmale werden in einem Schritt 102 zusammen mit der  
Bildsequenz oder in einer getrennten Datenbank abgespeichert.  
Mit Datenbank ist hier eine allgemeine Ansammlung von Daten  
gemeint, die optional mit geeigneten Zugriffsmechanismen  
20 versehen ist. Solche Datenbanken sind z.B. als funktionales  
Programmierinterface (z.B. als "application programming  
interface" API) oder als fertige, eigenständig auf einem  
Rechner laufende Programme dem Fachmann hinlänglich bekannt.

25 Über den Zugriff 103 werden die Suchmerkmale mit einer  
vorgegebenen Information verglichen und, falls eine  
Übereinstimmung gefunden wird, zu der Bildfolge, die durch  
das übereinstimmende Suchmerkmal referenziert (mit einem  
Verweis wird auf die Bildfolge gezeigt) wird, gesprungen  
30 werden kann. Wird keine Übereinstimmung zwischen der  
vorgegebenen Information und den Suchmerkmalen gefunden, so  
wird vorzugsweise dies dem Benutzer angezeigt.

In Fig.2 ist über den Verlauf einer Zeitachse  $t$  eine  
35 Bildsequenz BS, die Bilddaten BD und Toninformation TI  
umfaßt, dargestellt. Innerhalb der Bildsequenz BS ist eine  
Bildfolge BF als Teil der Bildsequenz BS gezeigt. In Fig.2

wird deutlich, daß die Bilddaten BD und die Toninformation TI eine zeitliche Entsprechung aufweisen, also jedem Tonausschnitt ein eindeutiger Bildausschnitt und umgekehrt zugeordnet werden kann.

5

Fig.3 zeigt die Bildsequenz BS, die mindestens eine Kombination aus einem Header H und Bilddaten BD (siehe beispielsweise die entsprechende Festlegung im MPEG-Standard: GOP = Group of Pictures) umfaßt, der Suchmerkmale M  
10 beispielhaft vorangestellt worden sind. Ebenso können die Suchmerkmale M der Bildsequenz BS angehängt oder innerhalb der Bildsequenz BS abgespeichert werden. Alternativ dazu ist es möglich, die Suchmerkmale M in einer externen Datenbank EDB abzuspeichern und so Suchmerkmale mehrerer Bildsequenzen  
15 BS in einer Datenbank zu sammeln. In Fig.3 sind Verweise PTR skizziert, die zeigen, daß zu jedem Suchmerkmal auch ein solcher Verweis gehört, der auf ein Bild innerhalb der Bildsequenz zeigt und somit eine Bildfolge in der Bildsequenz dereferenziert (d.h. die Bildfolge ist durch den Verweis PTR  
20 adressierbar).

Eine Organisation einer Liste von Suchmerkmalen und deren Beziehung zu den Bilddaten BD ist in Fig.4 dargestellt. Es gilt wieder der vertikale Verlauf einer Zeitachse t von oben  
25 nach unten. Die Suchmerkmale M sind in Form einer Liste der Bildsequenz BS, die den Header H und die Bilddaten BD enthält, vorangestellt. Die Liste enthält mehrere Suchmerkmale SM1 und SM2. Jedes Suchmerkmal umfaßt jeweils eine beschreibende Komponente INFO1 und INFO2 und jeweils  
30 einen Verweis (Zeiger, engl.: pointer) PTR1 und PTR2 auf ein Bild innerhalb der Bilddaten BD.

Eine zentrale Komponente in Fig.5 ist der Spracherkenner SE. Dort wird die Toninformation TI bearbeitet.

35

Eine Möglichkeit ist die Erkennung der in der Toninformation enthaltenen Sprache mit den die eine jeweilige Sprache



kennzeichnenden Worte. Diese Worte stellen Suchmerkmale  $SM_i$  ( $i=1,2,3,\dots,n$ ) dar, die in die Liste mit Suchmerkmalen LSM eingetragen werden.

- 5 Aus der Toninformation erkannte Worte können der Häufigkeit nach sortiert dem Benutzer dargestellt werden, so daß dieser eine Auswahlmöglichkeit für die vorgebbare Information hat.

- Auf Wortebene können geeignete Filter nur bestimmte
- 10 Wortarten, z.B. Substantive, zum Abspeichern oder nur bestimmte vorgegebene Worte, die in einem anwendungsabhängigen Lexikon abgelegt werden, zulassen. Im zweiten Fall kann man gezielt die Toninformation nach Kategorien einstufen. Jede Kategorie umfaßt ein bestimmtes
- 15 anwendungsabhängiges Lexikon. Ein Beispiel für ein anwendungsabhängiges Lexikon ist ein Sportlexikon mit Worten, die in dieser Themenklasse eine Rolle spielen. Die Toninformation einer Bildsequenz wird auf Übereinstimmungen mit in diesem Sportlexikon vorhandenen Worten untersucht.
- 20 Jede Übereinstimmung führt zu einem Suchmerkmal, also einer beschreibenden Komponente INFO und einem Verweis PTR auf das Bild, zu dem das jeweilige Wort aufgetreten ist.

- Auch ist es möglich, nicht die der Sprache eigenen Worte als
- 25 Suchmerkmale  $SM_i$ , sondern Laute, d.h. Phoneme bzw. Phonemkombinationen, abzuspeichern und somit von einer speziellen Sprache unabhängig zu sein.

- Bei der Suche nach einer Bildfolge wird die vorgegebene
- 30 Information, nach der gesucht werden soll, eingegeben. Diese Eingabe kann geschehen auf Wortebene, z.B. mittels einer Tastatur oder in Form gesprochener Sprache über ein Mikrofon MIK.

- 35 Es kann auch direkt über das Mikrofon MIK oder über eine andere Eingabeeinheit, z.B. eine Tastatur, die vorgegebenen Information eingegeben werden und ohne vorhandene

abgespeicherte Suchmerkmale S<sub>Mi</sub> die zu der Bildszene  
gehörende Toninformation TI nach der vorgegebenen Information  
abgesucht werden (siehe Verknüpfung 501). Dabei kann auf  
Wortebene einer jeweiligen Sprache oder auf Lautebene  
5 (Phoneme/Phonemkombinationen) nach der vorgegebenen  
Information innerhalb der Toninformation TI gesucht werden.

Im Rahmen dieses Dokuments wurde folgende Veröffentlichung zitiert:

- [1] Schukat-Talamazzini: Automatische Spracherkennung, Vieweg-Verlag, 1995.

Patentansprüche

1. Verfahren zum Abspeichern von Suchmerkmalen einer  
Bildsequenz, die Toninformation umfaßt, und Zugriff auf  
5 eine Bildfolge, die eine echte Teilmenge der Bildsequenz  
ist,
  - a) bei dem die Suchmerkmale aus der Toninformation  
ermittelt werden,
  - b) bei dem die Suchmerkmale abgespeichert werden,
  - 10 c) bei dem die Suchmerkmale mit einer vorgegebenen  
Information verglichen werden und,  
falls eine übereinstimmendes Suchmerkmal gefunden  
wird, zu der Bildfolge, die mit dem übereinstimmenden  
Suchmerkmal verknüpft ist, gesprungen wird,
  - 15 oder eine Ausgabe, daß keine Übereinstimmung  
zwischen vorgegebener Information mit den  
Suchmerkmalen gefunden worden ist, dargestellt wird.
2. Verfahren nach Anspruch 1,  
20 bei dem durch ein Spracherkennungssystem vorgebbare Worte  
aus der Toninformation ausgewertet, erkannt und  
abgespeichert werden.
3. Verfahren nach Anspruch 2,  
25 bei dem die vorgebbaren Worte Substantive sind.
4. Verfahren nach einem der Ansprüche 1 bis 3,  
bei dem die Suchmerkmale gemeinsam mit der Bildsequenz  
abgespeichert werden.  
30
5. Verfahren nach einem der Ansprüche 1 bis 4,  
bei dem die Suchmerkmale in einer von der Bildsequenz  
separaten Datenbank abgespeichert werden.
- 35 6. Verfahren nach einem der Ansprüche 1 bis 5,  
bei dem die vorgegebene Information mittels gesprochener  
Sprache eingegeben wird.

7. Verfahren nach einem der Ansprüche 1 bis 6,  
bei dem zu der Bildsequenz eine Liste mit den häufigsten  
aus der Toninformation zu dieser Bildsequenz erkannten  
5 Worten angezeigt wird.
8. Verfahren nach einem der Ansprüche 1 bis 7,  
bei dem als Suchmerkmale Phoneme und/oder  
Phonemkombinationen abgespeichert werden.
- 10 9. Verfahren zur Suche nach einer Sprache umfassenden  
vorgegebenen Information in einer Bildsequenz, die  
Toninformation umfaßt, und Zugriff auf eine Bildfolge,  
die eine echte Teilmenge der Bildsequenz ist,
- 15 a) bei dem die vorgegebene Information in der  
Toninformation der Bildsequenz gesucht wird,  
b) bei dem die Bildfolge ab einem Bild, dessen  
Toninformation die vorgegebene Information enthält,  
angezeigt wird.
- 20 10. Verfahren nach Anspruch 9,  
bei dem durch ein Spracherkennungssystem vorgebbare Worte  
aus der Toninformation erkannt werden.
- 25 11. Verfahren nach Anspruch 9 oder 10,  
bei dem die vorgegebene Information mittels gesprochener  
Sprache eingegeben wird und durch das  
Spracherkennungssystem erkannt wird.
- 30 12. Verfahren nach einem der Ansprüche 9 bis 11,  
bei dem Phoneme und/oder Phonemkombinationen der  
vorgegebenen Information mit Phonemen und/oder  
Phonemkombinationen aus der Toninformation verglichen  
werden.

1/3

FIG 1

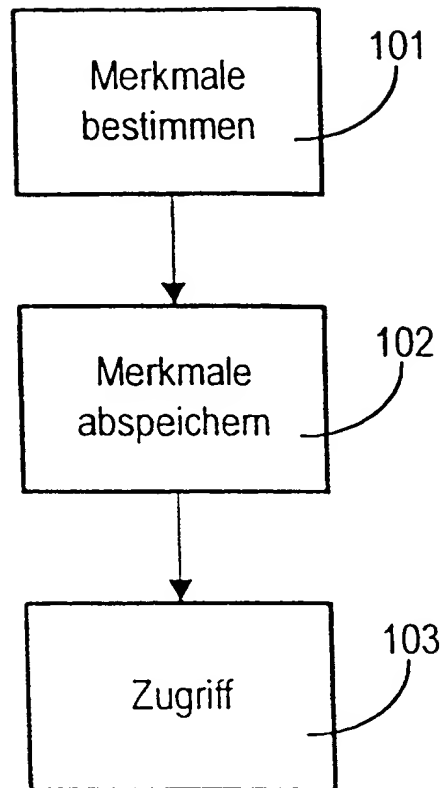
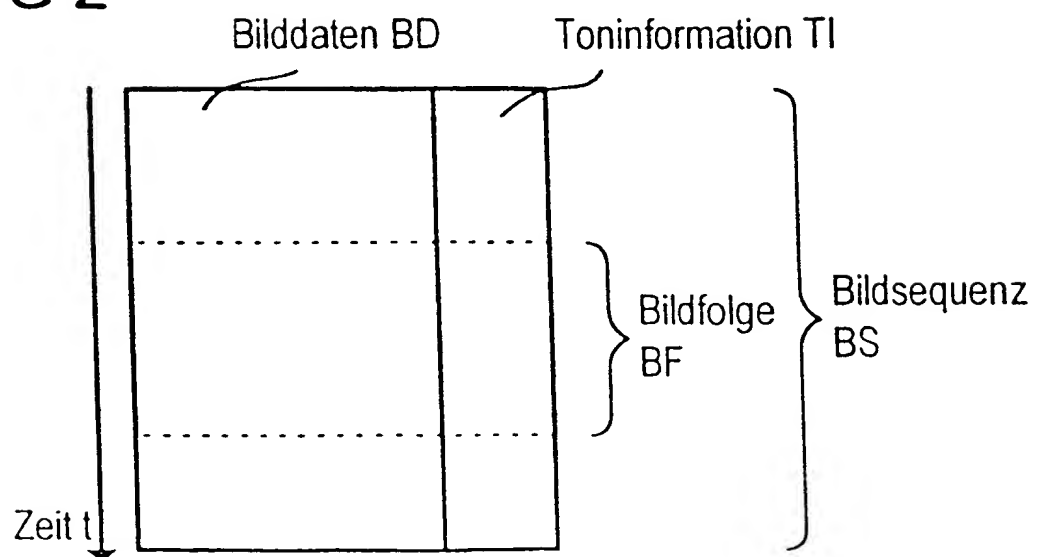


FIG 2



2/3

FIG 3

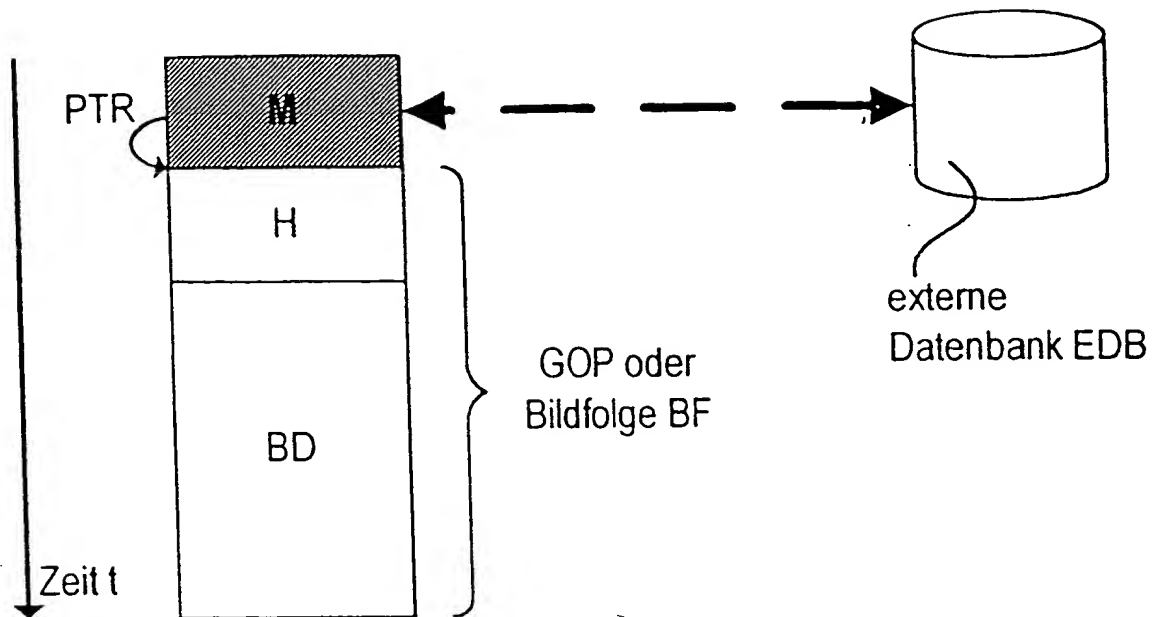
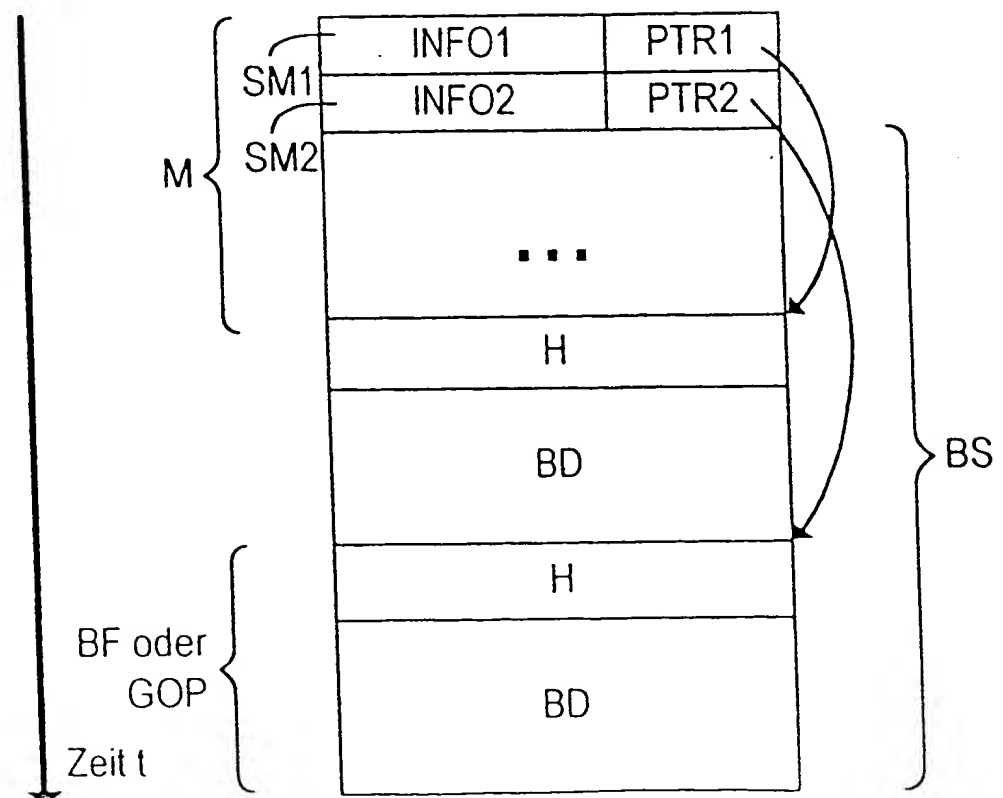
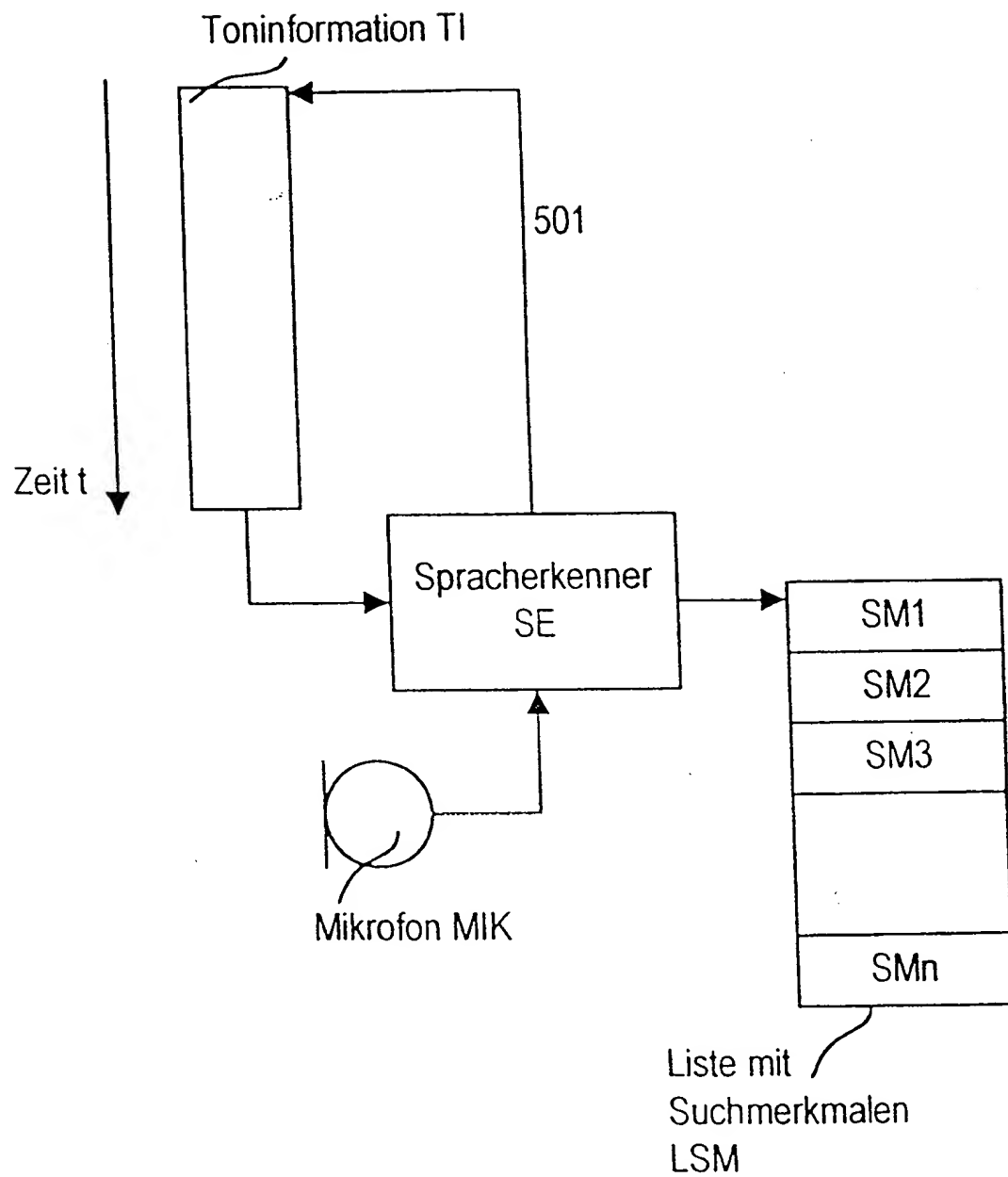


FIG 4



3/3

FIG 5





## INTERNATIONAL SEARCH REPORT

International Application No

PCT/DE 98/01985

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 G11B27/28 G11B27/10 G06F17/30 G10L5/06

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 G11B G06F G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	WACTLAR H D ET AL: "AUTOMATED VIDEO INDEXING OF VERY LARGE VIDEO LIBRARIES" SMPTE JOURNAL, vol. 106, no. 8, August 1997, pages 524-528, XP000698604 see the whole document X: Presentiert auf "138th SMPTE Technical Conference" . October 8 bis 12, 1996 ---	1-3, 5, 8-10, 12
X	"CORRELATING AUDIO AND MOVING-IMAGE TRACKS" IBM TECHNICAL DISCLOSURE BULLETIN, ISSN 0018-8689, vol. 33, no. 10A, 1 March 1991, page 295/296 XP000110048 USA see the whole document ---	1-3, 5, 9, 10



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

## \* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&amp;" document member of the same patent family

Date of the actual completion of the international search

11 January 1999

Date of mailing of the international search report

18/01/1999

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040. Tx. 31 651 epo nl.  
Fax: (+31-70) 340-3016

Authorized officer

Daalmans, F

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/DE 98/01985

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 507 743 A (STENOGRAPH CORP) 7 October 1992 see page 2, line 28 - line 46 see page 3, line 7 - page 4, line 19 see page 5, line 31 - line 38; claims 1,5,7,9,10 ---	1,4,9
X	HAUPTMANN A G: "SPEECH RECOGNITION IN THE INFORMEDIA DIGITAL VIDEO LIBRARY: USES AND ELIMINATIONS" PROCEEDINGS OF THE 7TH. INTERNATIONAL CONFERENCE ON TOOLS WITH ARTIFICIAL INTELLIGENCE, HERNDON, VA., NOV. 5 - 8, 1995, no. CONF. 7, 5 November 1995, pages 288-294, XP000598376 INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS see the whole document ---	1,2,5-7, 9-11
X	US 5 649 060 A (ELLOZY HAMED A ET AL) 15 July 1997 see column 3, line 32 - column 11, line 10 ---	1,2,5, 8-10,12
X	US 5 136 655 A (BRONSON BARRY S) 4 August 1992 see column 1, line 37 - line 62 see column 2, line 45 - column 4, line 5 ---	1,2,5,9, 10
A	EP 0 780 777 A (HEWLETT PACKARD CO) 25 June 1997 see the whole document ---	1,3,5,9
A	DE 32 16 871 A (HERTZ INST HEINRICH) 3 November 1983 see the whole document ---	
A	XUEDONG HUANG ET AL: "THE SPHINX-II SPEECH RECOGNITION SYSTEM: AN OVERVIEW" COMPUTER SPEECH AND LANGUAGE, vol. 7, no. 2, 1 April 1993, pages 137-148, XP000382200 LONDON, GB see the whole document ---	1,9

-/--

# INTERNATIONAL SEARCH REPORT

Int. l. Application No

PCT/DE 98/01985

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>SHANDLE J: "DIGITAL AUDIO DELIVERS  NON-STOP INNOVATIONS SPEECH AND MUSIC KEEP  PUSHING THE FRONTIERS OF COMPUTING ACROSS  TECHNOLOGIES INCLUDING COMPRESSION,  SYNTHESIS, AND SPEECH RECOGNITION"  ELECTRONIC DESIGN,  vol. 41, no. 22, 1 November 1993, pages  53/54, 56-58, 60, 64 - 66, XP000408631  CLEVELAND. OH. US  see 64 right column: "Speech  Recognition"- see 65, right column,  line 18</p> <p style="text-align: center;">-----</p>	

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/DE 98/01985

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
EP 0507743	A	07-10-1992	AU	1403092 A	08-10-1992
US 5649060	A	15-07-1997	EP	0649144 A	19-04-1995
			JP	7199379 A	04-08-1995
US 5136655	A	04-08-1992	NONE		
EP 0780777	A	25-06-1997	US	5794249 A	11-08-1998
			JP	9185879 A	15-07-1997
DE 3216871	A	03-11-1983	NONE		

# INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen

PCT/DE 98/01985

## A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES

IPK 6 G11B27/28 G11B27/10 G06F17/30 G10L5/06

Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK

## B. RECHERCHIERTE GEBIETE

Recherchierte Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)

IPK 6 G11B G06F G10L

Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

## C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
P, X	WACTLAR H D ET AL: "AUTOMATED VIDEO INDEXING OF VERY LARGE VIDEO LIBRARIES" SMPTE JOURNAL, Bd. 106, Nr. 8, August 1997, Seiten 524-528, XP000698604 siehe das ganze Dokument X: Präsentiert auf "138th SMPTE Technical Conference", Oktober 8 bis 12, 1996 ---	1-3, 5, 8-10, 12
X	"CORRELATING AUDIO AND MOVING-IMAGE TRACKS" IBM TECHNICAL DISCLOSURE BULLETIN, ISSN 0018-8689, Bd. 33, Nr. 10A, 1. März 1991, Seite 295/296 XP000110048 USA siehe das ganze Dokument ---	1-3, 5, 9, 10
	--- -/-	



Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen



Siehe Anhang Patentfamilie

\* Besondere Kategorien von angegebenen Veröffentlichungen :

- "A" Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist
- "E" älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist
- "L" Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)
- "O" Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht
- "P" Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist

- "T" Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist
- "X" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfindenscher Tätigkeit beruhend betrachtet werden
- "Y" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfindenscher Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist
- "&" Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche

11. Januar 1999

Absenddatum des internationalen Recherchenberichts

18/01/1999

Name und Postanschrift der Internationalen Recherchenbehörde

Europäisches Patentamt, P.B. 5818 Patentlaan 2  
NL - 2230 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax (+31-70) 340-2040

Bevollmächtigter Beauftragter

Doelmann, F.

## INTERNATIONALER RECHERCHENBERICHT

Int. ationales Aktenzeichen

PCT/DE 98/01985

C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	EP 0 507 743 A (STENOGRAPH CORP) 7. Oktober 1992 siehe Seite 2, Zeile 28 - Zeile 46 siehe Seite 3, Zeile 7 - Seite 4, Zeile 19 siehe Seite 5, Zeile 31 - Zeile 38; Ansprüche 1,5,7,9,10 ---	1,4,9
X	HAUPTMANN A G: "SPEECH RECOGNITION IN THE INFORMEDIA DIGITAL VIDEO LIBRARY: USES AND LIMINATIONS" PROCEEDINGS OF THE 7TH. INTERNATIONAL CONFERENCE ON TOOLS WITH ARTIFICIAL INTELLIGENCE, HERNDON, VA., NOV. 5 - 8, 1995, Nr. CONF. 7, 5. November 1995, Seiten 288-294, XP000598376 INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS siehe das ganze Dokument ---	1,2,5-7, 9-11
X	US 5 649 060 A (ELLOZY HAMED A ET AL) 15. Juli 1997 siehe Spalte 3, Zeile 32 - Spalte 11, Zeile 10 ---	1,2,5, 8-10,12
X	US 5 136 655 A (BRONSON BARRY S) 4. August 1992 siehe Spalte 1, Zeile 37 - Zeile 62 siehe Spalte 2, Zeile 45 - Spalte 4, Zeile 5 ---	1,2,5,9, 10
A	EP 0 780 777 A (HEWLETT PACKARD CO) 25. Juni 1997 siehe das ganze Dokument ---	1,3,5,9
A	DE 32 16 871 A (HERTZ INST HEINRICH) 3. November 1983 siehe das ganze Dokument ---	
A	XUEDONG HUANG ET AL: "THE SPHINX-II SPEECH RECOGNITION SYSTEM: AN OVERVIEW" COMPUTER SPEECH AND LANGUAGE, Bd. 7, Nr. 2, 1. April 1993, Seiten 137-148, XP000382200 LONDON, GB siehe das ganze Dokument ---	1,9
	---	
	-/--	

## C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	<p>SHANDLE J: "DIGITAL AUDIO DELIVERS NON-STOP INNOVATIONS SPEECH AND MUSIC KEEP PUSHING THE FRONTIERS OF COMPUTING ACROSS TECHNOLOGIES INCLUDING COMPRESSION, SYNTHESIS, AND SPEECH RECOGNITION" ELECTRONIC DESIGN, Bd. 41, Nr. 22, 1. November 1993, Seiten 53/54, 56-58, 60, 64 - 66, XP000408631 CLEVELAND, OH, US Seite 64, rechte Spalte: "Speech Recognition"- Seite 65, rechte Spalte, Linie 18</p> <p>-----</p>	

# INTERNATIONALER RECHERCHENBERICHT

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

Internationales Aktenzeichen

PCT/DE 98/01985

Im Recherchenbericht angeführtes Patentdokument		Datum der Veröffentlichung	Mitglied(er) der Patentfamilie		Datum der Veröffentlichung
EP 0507743	A	07-10-1992	AU	1403092 A	08-10-1992
US 5649060	A	15-07-1997	EP	0649144 A	19-04-1995
			JP	7199379 A	04-08-1995
US 5136655	A	04-08-1992	KEINE		
EP 0780777	A	25-06-1997	US	5794249 A	11-08-1998
			JP	9185879 A	15-07-1997
DE 3216871	A	03-11-1983	KEINE		